Project Report

|  |  |
| --- | --- |
| Product Name | Implement Data Visualisation techniques using MS Power BI |
| Qualification Name (NICF) | NICF Diploma in Infocomm Technology(Data) |
| Product Name | NICF-Data Queries and Visualization Basics(SF) |
| Module Name (NICF) | **NICF-**Data Queries and Visualization Basics**(SF)** |

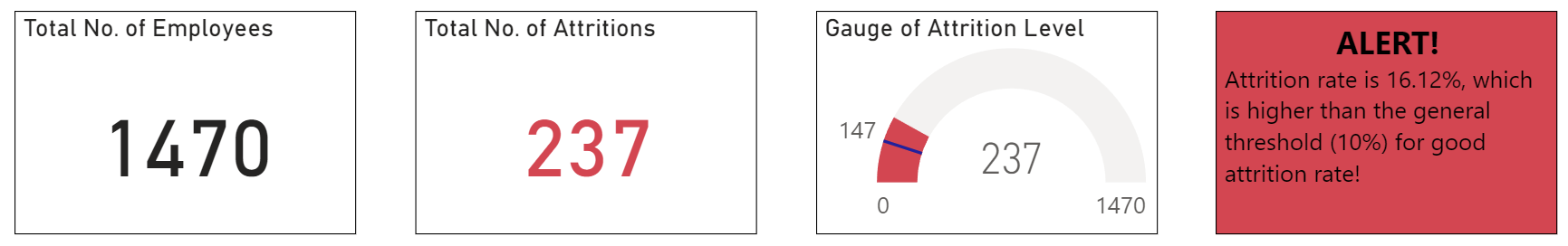
|  |  |  |  |
| --- | --- | --- | --- |
| Student name | | Assessor name | |
| Lee Jack Shiang | | Harsh Shah | |
| Date issued | Completion date | | Submitted on |
| 31-Mar-2022 | 11-Apr-2022 | | 11-Apr-2022 |
|  | |  | |
| Project title | HR Attrition Project | | |

|  |
| --- |
| Learner declaration |
| I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.  LEE JACK SHIANG  Student signature: Date: 11-Apr-2022 |

Content

1. Project background 2
2. Project Objectives 3
3. Project Specifications 4
4. Activity 1 5
5. Activity 2 8
6. Activity 3 9
7. Activity 4 10
8. Annexure (Excel Workbook) 19

Project Background (Explain the Project in your own words in 15 – 20 lines)

****

The background of this project is to investigate the attrition of the company, which is found to be at an alerting level (16.12%, which is higher than the general threshold at around 10% for good attrition rate). The data used for this project consists of 1471 rows (column header + 1470 employees) and 34 columns (attributes) and collected in the past 12 months. In the data set, several key information of the employees, for example, their demographics (gender, age etc.), department, income level, their feedback on relationships and work-life balance etc., will be used for study of the attrition.

General speaking, good attrition rate (less than 10%) can be beneficial to the company as those new employees can be more qualified, capable, and productive than his or her predecessors. Also, these new employees may bring new or innovative ideas which may have positive impact to the company. However, if the attrition rate is higher than certain threshold (about 10%), it may have negative impact to the company in the long run, as the hiring process for new employees can be expensive, time consuming, and difficult to find truly qualified candidates. As such, it is important for the company to find out the reasons why employees quit and identify those main factors (ie. major contributors for its high attrition).

Below are some of the reasons why employees quit:

* Poor relationships with colleagues
* Poor working environment
* Seeking higher salary
* Looking for better work-life balance
* Feeling uninspired in the current job
* For better job growth and career advancement

Project Objective

The 3 main objectives of this project are:

1. Compile a report using Power BI Services and its visualization tools to provide meaningful insights from the raw data set
2. Find out those main factors for the attrition
3. Provide recommendations to the company with the aim of reducing attrition back to healthy state (<10%)
4. Project Specifications

Technical tools used in this project are:

1. MS Excel – For data processing and data import into Power BI services
2. Power BI Services – For development and sharing of visualization report
3. MS Word – For compilation of final report

Main steps for tackling this project:

1. Data acquisition
   * The dataset and data definition files (MS Excel) shall be downloaded from LMS for this project
2. Data processing
   * The original dataset contains numerical values for some attribute columns, and may not be as interpretable if short text descriptions were used
   * The numerical values in those columns shall be replaced by the corresponding short text descriptions from the data definition file for better interpretability of dataset and facilitate subsequent data visualization stage
3. Data importing into Power BI Services
   * The processed dataset shall be imported into Power BI Services and ready for the development of visualizations
4. Grouping for factors
   * Before the commencement of building any visualizations, those related attributes shall be grouped together to have more systematic visualizations. For example, grouping demographics attributes like gender, age, marital status etc. together into a single page (tab)
   * For this project, the visualizations are based on the following groups:
     + Demographics
     + Department
     + Relationship, history with company
     + Income, performance, and salary hike
     + Distance from home and business travel
     + Work-life balance and welfare
5. Development of visualizations in Power BI Services
   * Based on the grouping above, the development of visualizations using Power BI Services splits into several pages (tabs) based on those groups
   * Several types of visualizations like donut chart, bar chart, tree map etc. shall be utilized where appropriate for finding insights
6. Finding insights from the visualizations
   * Based on the visualizations developed above, some insights should be found, and they should be noted down in their respective report pages for side-by-side reference with the visualizations
   * Based on insights found from those visualizations, some recommendations shall be provided to the company with the aim of reducing the attrition back to healthy state (<10%)
7. Compilation of final report
   * The final MS Word report shall be compiled as per the template provided in LMS for formal documentation of this project
8. Activity 1

Identify the factors influencing the Data Visualization:

* Extract data from HR Attrition excel
  + The data used for this project consists of 1471 rows (column header + 1470 employees) and 34 columns (attributes)
  + In the data set, several key information of the employees, for example, their demographics (gender, age etc.), department, income level, their feedback on relationships and work-life balance are provided
  + The 34 columns (attributes) of the data set are listed in the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age | AgeGroup | Attrition | BusinessTravel | DailyRate |
| Department | DistanceFromHome | Education | EducationField | Employee Count |
| Environment Satisfaction | Gender | HourlyRate | JobInvolvement | JobLevel |
| JobRole | JobSatisfaction | MaritalStatus | MonthlyIncome | MonthlyRate |
| NumCompanies Worked | OverTime | PercentSalaryHike | Performance Rating | Relationship Satisfaction |
| Standard Hours | StockOptionLevel | TotalWorkingYears | TrainingTimes LastYear | WorkLifeBalance |
| YearsAtCompany | YearsInCurrentRole | YearsSinceLast Promotion | YearsWith CurrManager |  |

* + As mentioned in the data processing step in Section 3, numerical values in some attribute columns shall be replaced by short text descriptions for better interpretability. As such, the numerical values of the following attribute columns shall be replaced as per the data definition file:
    - Attrition: 0 ‘No’, 1 ‘Yes’
    - Business Travel: 0 ‘Non-Travel’, 1 ‘ Travel\_Rarely’, 2 ‘Travel\_Frequently’
    - Department: 1 ‘HR’, 2 ‘R&D’, 3 ‘Sales’
    - Education: 1 ‘Below College’, 2 ‘College’, 3 ‘Bachelor’, 4 ‘Master’, 5 ‘Doctor’
    - Environment Satisfaction: 1 ‘Low’, 2 ‘Medium’, 3 ‘High’, 4 ‘Very High’
    - Job Involvement: 1 ‘Low’, 2 ‘Medium’, 3 ‘High’, 4 ‘Very High’
    - Job Satisfaction: 1 ‘Low’, 2 ‘Medium’, 3 ‘High’, 4 ‘Very High’
    - Marital Status: 1 ‘Single’, 2 ‘Married’, 3 ‘Divorced’
    - Overtime: 0 ‘No, 1 ‘Yes’
    - Performance Rating: 1 ‘Low’, 2 ‘Good’, 3 ‘Excellent’, 4 ‘Outstanding’
    - Relationship Satisfaction: 1 ‘Low’, 2 ‘Medium’, 3 ‘High’, 4 ‘Very High’
    - Work-life Balance: 1 ‘Bad’, 2 ‘Good’, 3 ‘Better’, 4 ‘Best’

Notes:

1. The data definition shall follow the order of priority below:

Sheet1 -> Sheet2 (ie. All the data definition shall be followed as per information in Sheet1. If any required information cannot be found in Sheet1, then information in Sheet2 shall be followed)

1. For Job Level, the numerical values are defined as follows:

1 ‘Entry-Level’, 2 ‘Intermediate’, 3 ‘First-Level Management’,

4 ‘Middle Management’, 5 ‘Senior Management’

* Identify which of the visualizations – static or interactive is required

Both static and interactive visualizations will be used.

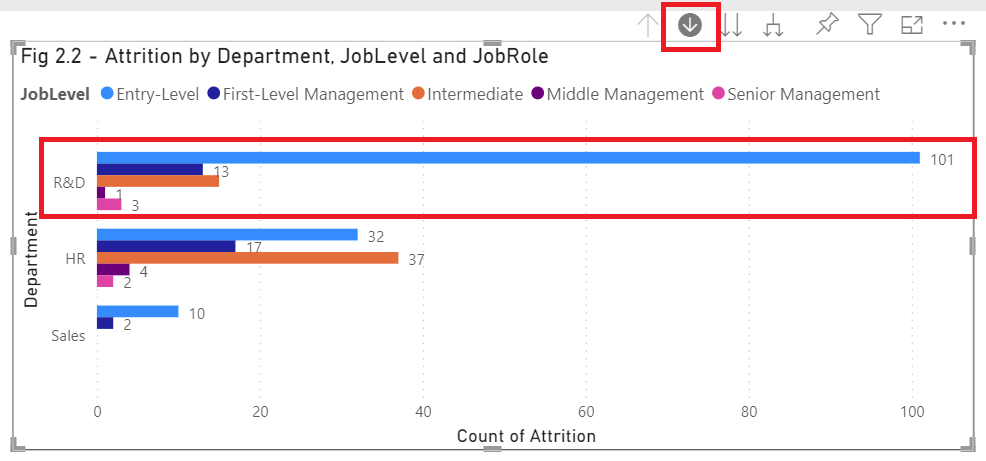
Example use case for static visualizations:

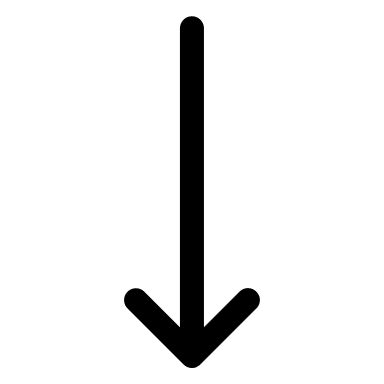
* + Text boxes can be utilized for noting down insights found in the visualizations.

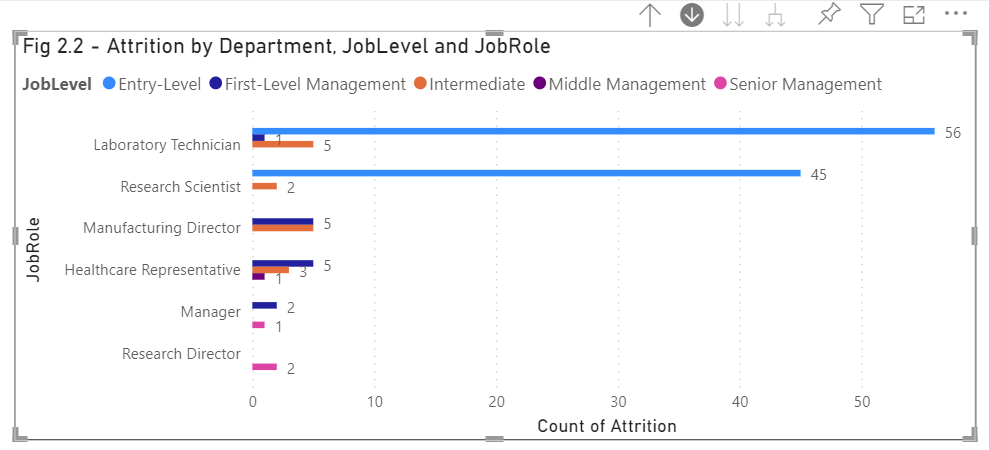
Example use case for interactive visualizations:

* + Bar chart, donut chart and tree maps etc. for which the visualizations will interact with the underlying dataset, shall be utilized for finding insights
* Identify the audience of data visualization and their size
  + Target audience: Senior Management of the company (Attrition = No)
  + Size: 64
* Identify the level of details required for data visualization (drill down) (grouping)
  + In the Power BI Services report, some of the visualizations can utilize the drill down function to show more detailed information of a particular visualization. Drill down can be used for some tree maps and bar charts in the report.
  + Example use cases:

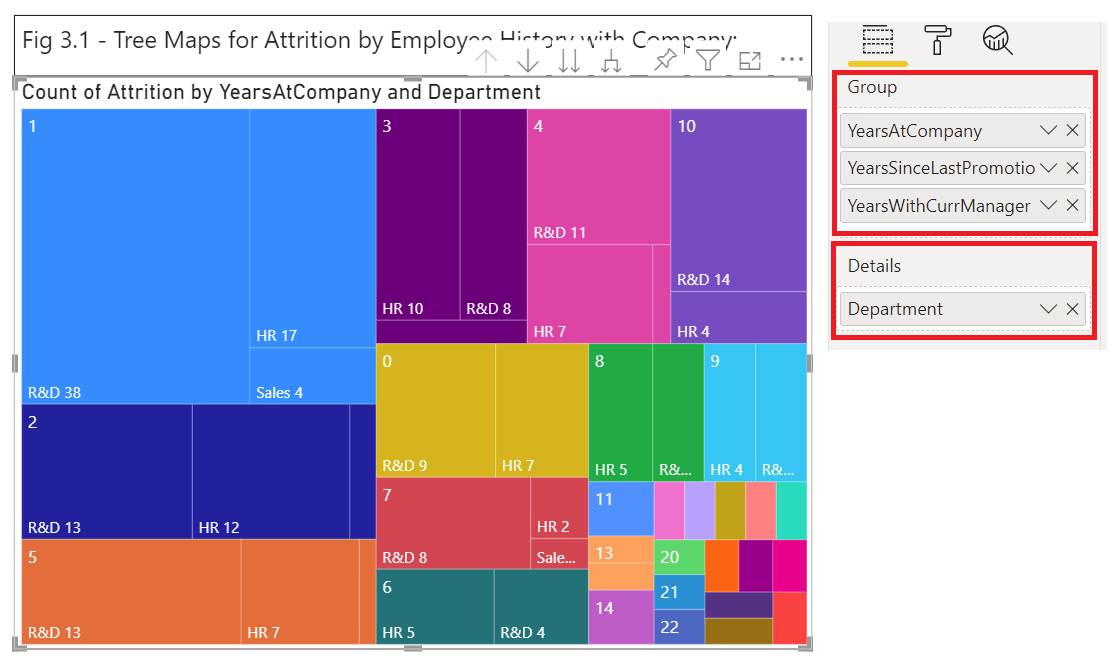
Drill down can be used for bar chart for additional details of the visualization. As shown in the figure below, drill down can be used for the R&D department for more detailed information about the attrition of this department. It is found that the main contributors of attrition of this department are from Entry-Level Laboratory Technician and Research Scientist.





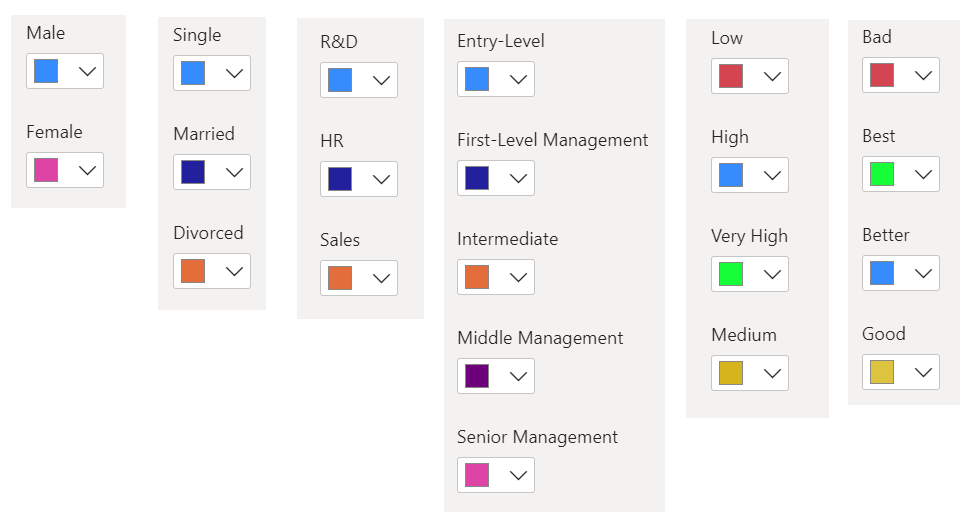


In addition, drill down can also be used for tree maps in the report to show attrition based on different groups, with details of showing the corresponding departments.



* Decide the colours required for visualization

Generally, colours used for data colours of visualizations follow the recommendations / defaults by Power BI Services. However, some colours should be changed for some of the visualizations where it would be more intuitive to use some other colours for presentation (eg. red colour for bad indication, green colour for good indication etc. Some examples for data colours are shown below:



* Decide the layout of the visualization

The report is prepared based on 6 groups (tabs) of attributes:

1. Demographics:

Gender, marital status, age group, education / education field

1. Department:

Department, job level, job role

1. Relationship, History with Company:

Years at company, years since last promotion, years with current manager, environment satisfaction, relationship satisfaction, job satisfaction

1. Income, Performance and Salary Hike:

Monthly income, performance rating, percent salary hike

1. Distance from Home and Business Travel:

Distance from home, frequency for business travel

1. Work-Life Balance and Welfare:

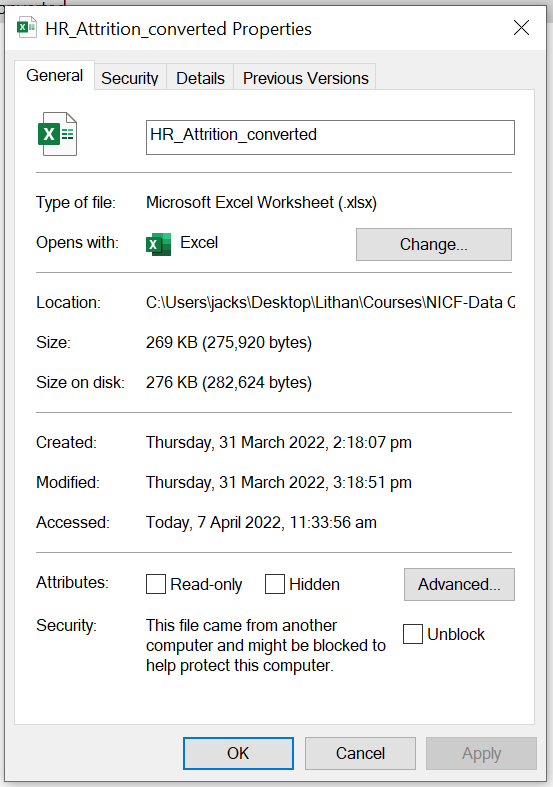
Standard hours, overtime, work-life balance, stock-option level

After the visualizations are created as per groups above, insights from the visualizations shall be noted down in each tab of the corresponding group. Finally, a summary tab for the insights found and recommendations will be provided.

1. Activity 2

Review the Project Data Set:

* Identify the total volume of data



The dataset is of MS excel type (.xlsx) with size of 276KB on disk.

Within the dataset it contains 1471 rows (column header + 1470 employees) and 34 columns for attributes.

* Analyze how diverse the data is based on job role, age group, education etc.

The following attributes are used for analyzing the diversity of the data:

* + Age group:
    - Age groups in the data range from 18 to 56 and above, which should cover all employees with legal ages working for the company
  + Department:
    - All 3 departments of the company are covered in the dataset
  + Education:
    - Employees with education qualifications from below college up to doctorate are considered in the dataset
  + Education field:
    - Employees with all education fields related to the company business are considered in the dataset
  + Gender:
    - Both male and female are considered in the dataset
  + Job level:
    - Job levels from entry level up to senior management are included in the dataset
  + Job role:
    - All job roles of the company are included in the dataset
  + Marital status:
    - All 3 marital status (single, married, divorced) are considered in the dataset

Based on above, it is concluded that the dataset should be diverse enough for analysis.

1. Activity 3

Select the appropriate data visualization technique for gathering the required insights from the Project data set:

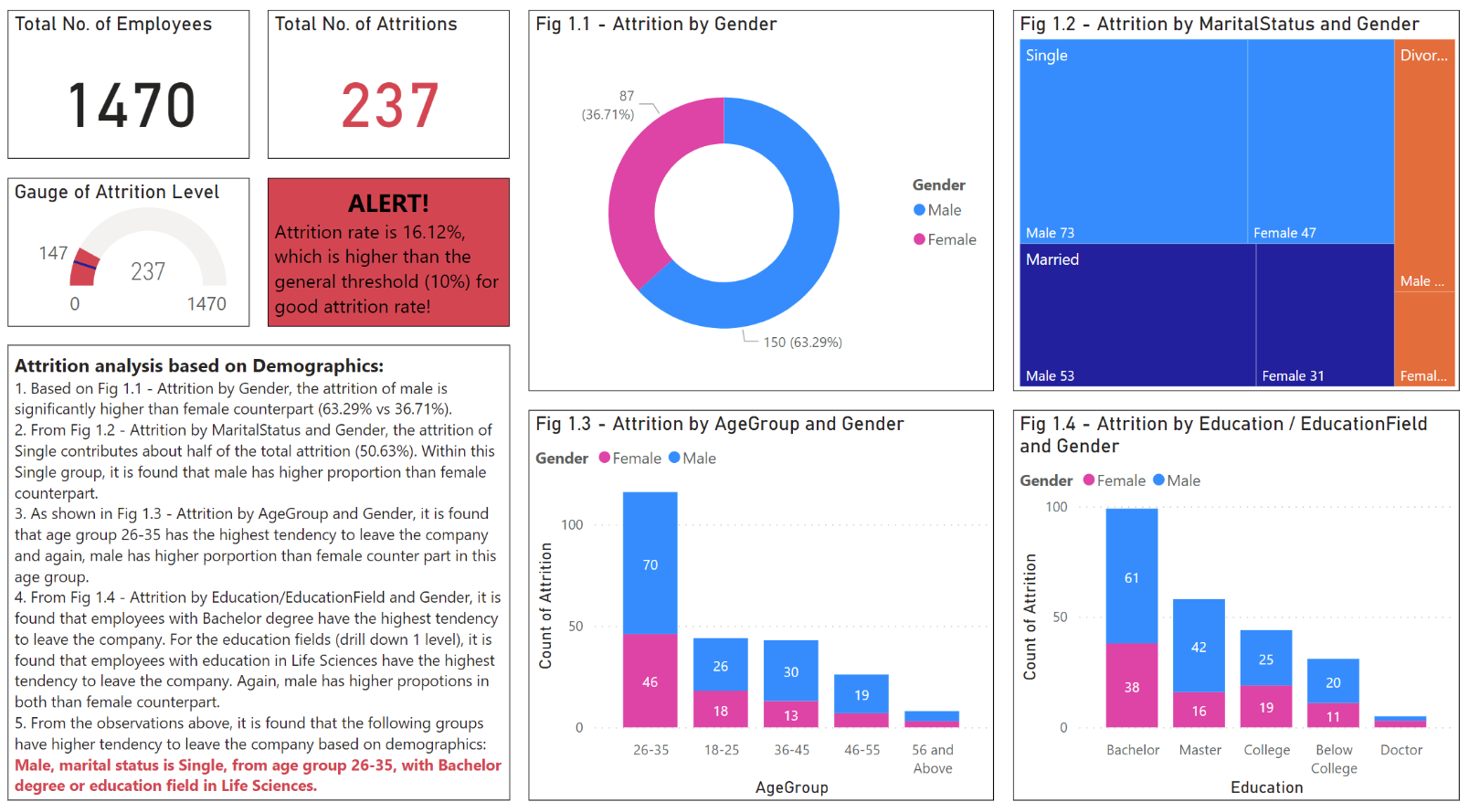
* Select one or more data visualization techniques among the following – line charts, word clouds, network diagram, bar chart, scatter plot, pie chart used in the Project
* Justify why you are using the techniques and why you have ruled out other techniques

|  |  |
| --- | --- |
| **Visualisation** | **Justification** |
| Bar / Column chart | It is an effective tool for comparing the values of different categories since the lengths of the bars are built proportional to the values that they represent, hence gives us a quick sense of how the values of categories compared to each other side-by-side |
| Card | It is normally used for displaying a single and critical number. In the report, card is used for displaying the total number of employees and total number of attritions in the first page of the report in order to quickly draw attention from the senior management |
| Donut chart | It is an effective tool for showing the proportion for value of each category. It gives us the value, as well as the percentage in the donut chart. However, it is advisable not to use donut chart if the number of categories is more than 5 |
| Gauge | It is a good choice if we are to compare the value of attribute to certain threshold. In the report, the gauge chart is used to compare the attrition no. to the general threshold for good attrition rate. Again, the gauge chart is put in the first page of the report to quickly draw attention from the senior management |
| Line and stacked  column chart | For quick comparison of 2 sets of attributes (eg. attrition and average monthly income by job role), line and stacked column chart is a solid choice since the 2 visuals (line and bars) are combined together in the same chart |
| Tree map | Tree map is a great choice when the number of categories of  attribute is large (eg. years at company), and for displaying large amount of hierarchical data |
| 100% stacked  column chart | It is a great choice for showing the relative percentage of multiple categories in stacked columns where the total of stacked columns always equals 100%. In the report, the 100% stacked column chart is used for showing the proportion (percentage) of attrition by work-life balance rating for each job role |
| Text box | Text box is used for displaying title or static information. In the report it is primarily used for narratives of insights found in the visualisations |

As the target audience is the senior management of the company, it is preferable to use relatively simple visualizations like card, gauge chart, donut chart and bar chart in order to quickly draw their attention since they may not be as technical or analytical. However, there may still be times where more complex visualizations like tree map are required, in order to properly display the intended visualizations.

1. Activity 4

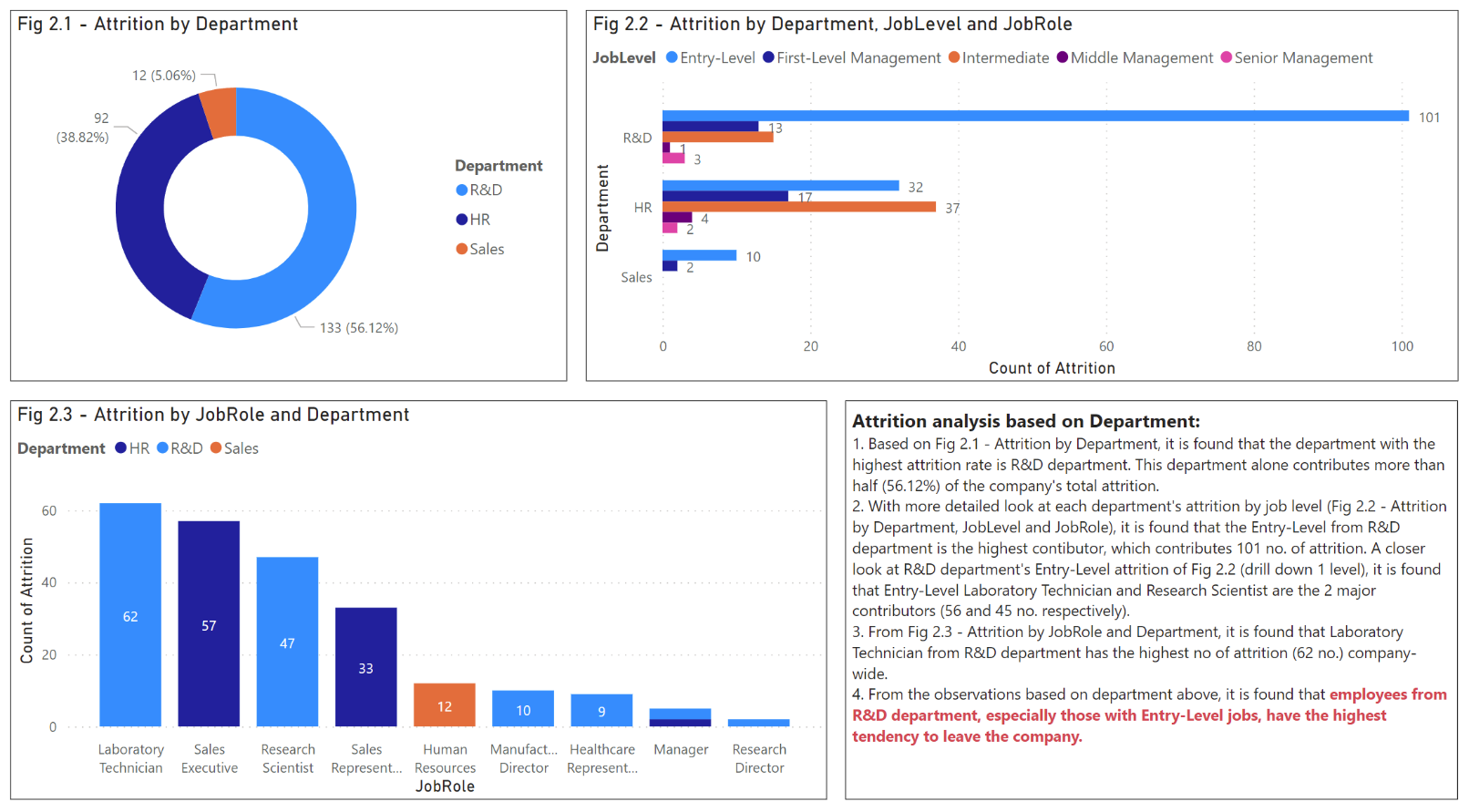
Page 1: Attrition by Demographics



|  |  |  |
| --- | --- | --- |
| **Visuals** | **Outcome** | **Justification** |
| Card:  Total No. of Employees  Total No. of Attritions | -Total no. of employees is 1470  -Total no of attritions is 237 | Card is useful for displaying single and critical number for quickly drawing attention of the audience |
| Gauge:  Gauge of Attrition Level | -Show the alarming level of attrition since it exceeds the general threshold for good attrition rate (<10%) | Gauge is useful for comparing the concerned figure with certain threshold at a glance |
| Donut Chart:  Fig 1.1 - Attrition by Gender | -Show the attrition by gender  -Male has higher attrition (63.29%) than female (36.71%) | Donut chart is a good choice for showing the proportion for value of each category. It gives us the value, as well as the percentage in the donut chart.  In Fig 1.1, only 2 categories for gender. Hence it is appropriate to use donut chart |
| Treemap:  Fig 1.2 – Attrition by MaritalStatus and Gender | -Show the attrition by marital status and gender  -Single contributes about half of the attrition (50.63%) and male has higher proportion than female | Tree map is a great choice when the number of categories of attribute is large, and for displaying large amount of hierarchical data  In Fig 1.2, since the 3 marital statuses (single, married and divorced) will be further divided based on gender (male and female) and this introduce some additional hierarchy, it is more appropriate to display the visualization using treemap |
| Stacked Column Chart:  Fig 1.3 – Attrition by AgeGroup and Gender | -Show the attrition by age group and gender  -Age group 26-35 has the highest attrition, and male has higher proportion than female | Stacked column chart work well when the focus of the chart is to compare the totals and one part of the totals  In Fig 1.3, it is used for investigating which age group has the highest attrition, and within which male or female has higher proportion |

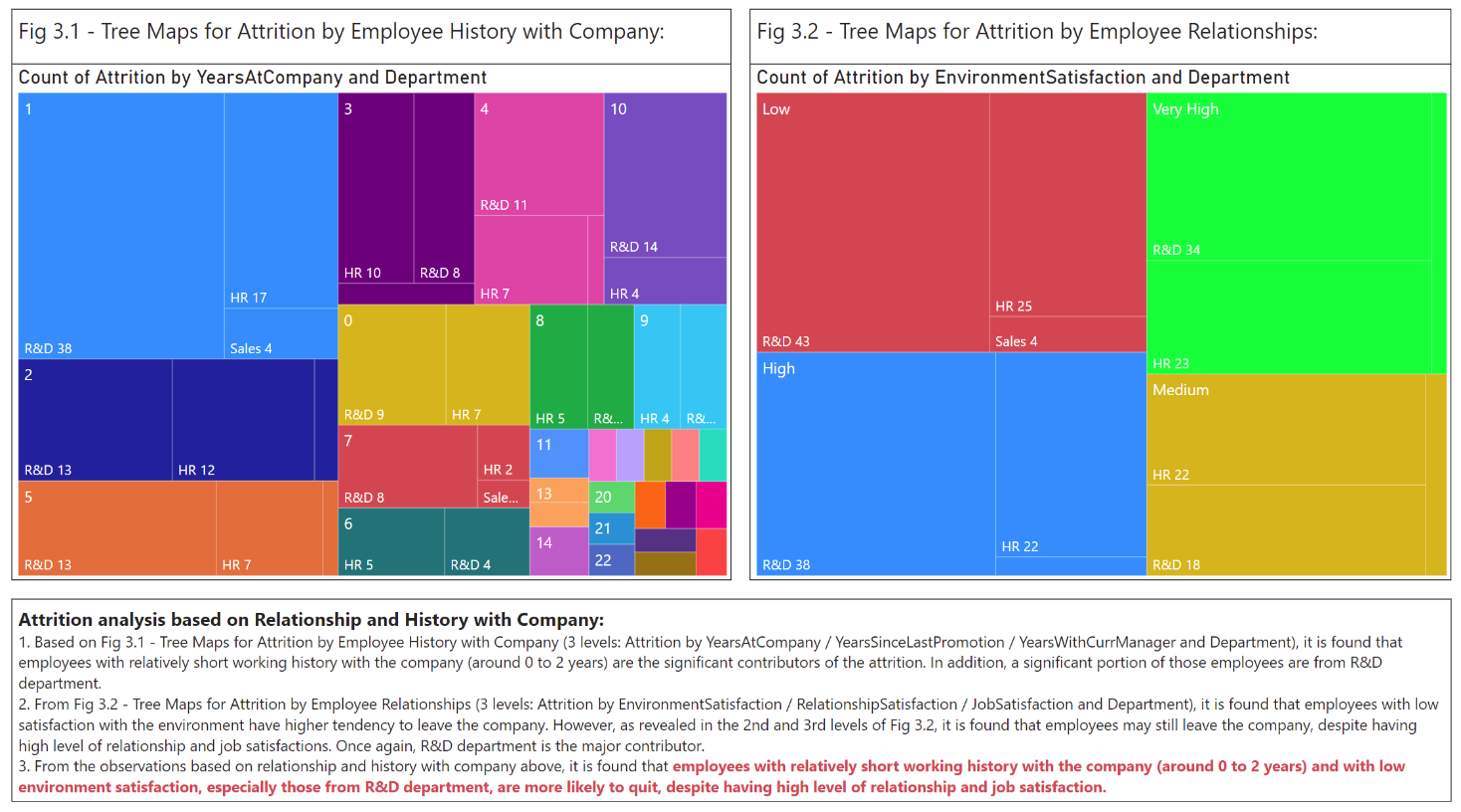
|  |  |  |
| --- | --- | --- |
| **Visuals** | **Outcome** | **Justification** |
| Stacked Column Chart:  Fig 1.4 – Attrition by Education / EducationField and Gender | -Show the attrition by education / education field and gender  -Employees with Bachelor degree has the highest attrition  -Employees with education field in Life Sciences has the highest attrition  -Male has higher proportion in the attrition by both education and education field | Stacked column chart work well when the focus of the chart is to compare the totals and one part of the totals  In Fig 1.4, it is used for investigating which education / education field groups have the highest attrition, and within which male or female has higher proportion |

Page 2: Attrition by Department



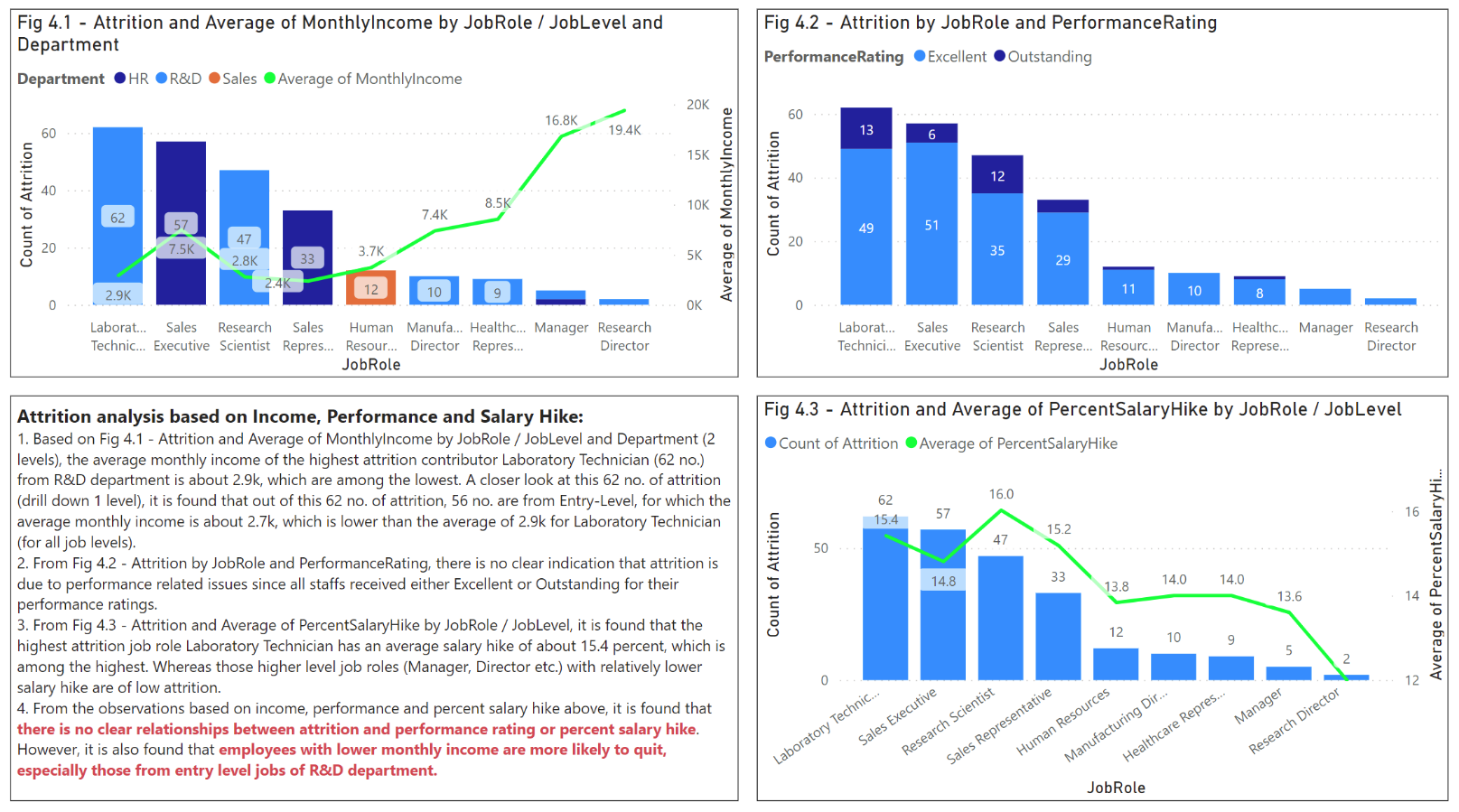
|  |  |  |
| --- | --- | --- |
| **Visuals** | **Outcome** | **Justification** |
| Donut Chart:  Fig 2.1 - Attrition by Department | -Show the attrition by department  -R&D department has the highest attrition (56.12%) | Donut chart is a good choice for showing the proportion for value of each category. It gives us the value, as well as the percentage in the donut chart.  In Fig 2.1, only 3 categories for department. Hence it is appropriate to use donut chart |
| Clustered Bar Chart:  Fig 2.2 – Attrition by Department, JobLevel and JobRole | -Show the attrition by department, job level and job role  -Entry level jobs from R&D department is the major contributor (101 no.)  -Drilling down 1 level for R&D department, it reveals that entry level Laboratory Technician (56 no.) and Research Scientist (45 no.) are the 2 main contributors | Clustered bar chart gives direct comparison of multiple subcategories in a given category and it is a good option for visualising the proportion of each subcategory compared to others  In Fig 2.2, attritions by subcategories (job level or job role) are shown for each category (department). By using clustered bar chart, it gives us a quick sense of which job level / job role from which department has the highest attrition |
| Stacked Column Chart:  Fig 2.3 – Attrition by JobRole and Department | -Show the attrition by job role and department  -Laboratory Technician from R&D department has the highest attrition (62 no.) | Stacked column chart work well when the focus of the chart is to compare the totals and one part of the totals  In Fig 2.3, it is used to compare the attrition of individual job role with each other |

Page 3: Attrition by Relationship, History with Company



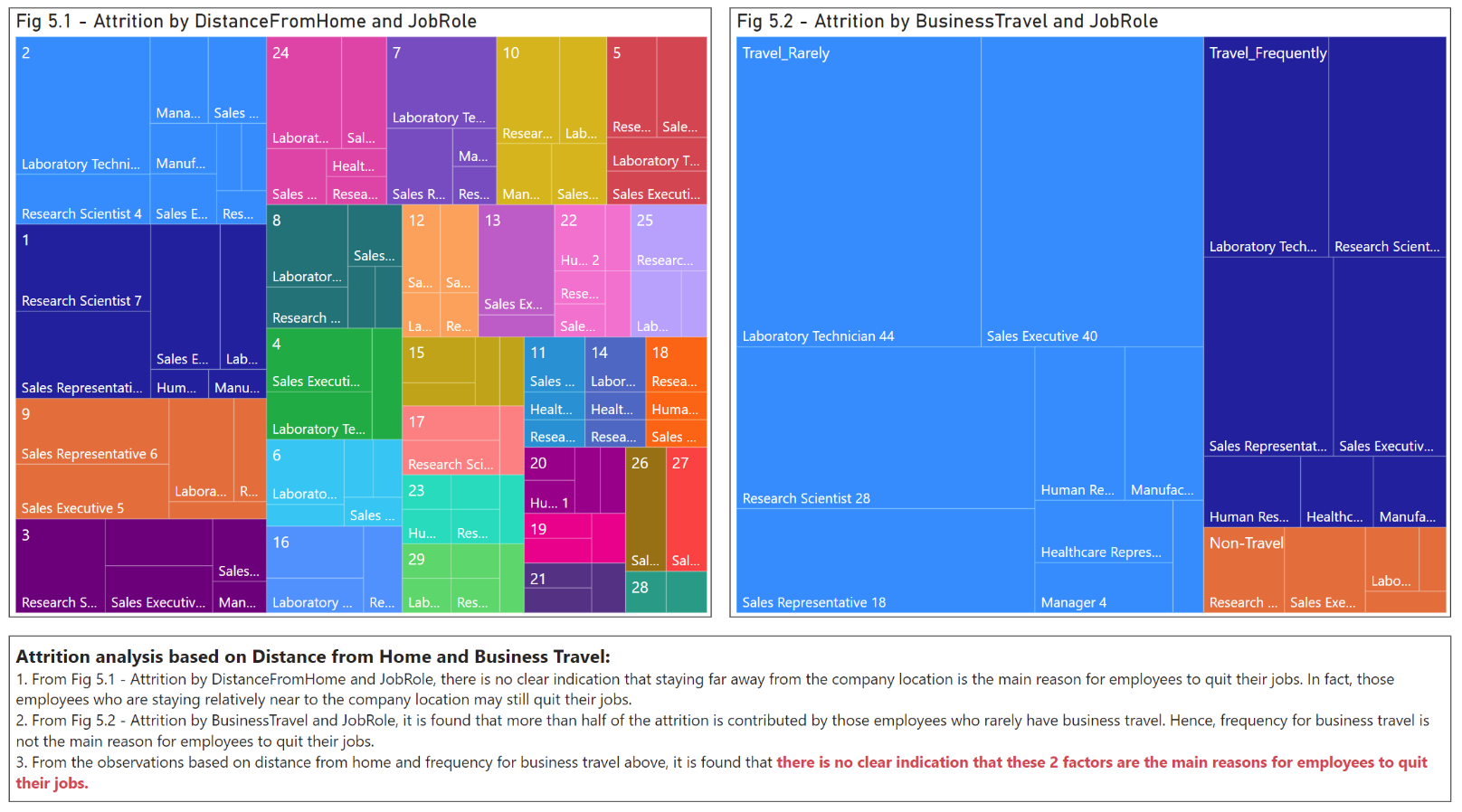
|  |  |  |
| --- | --- | --- |
| **Visuals** | **Outcome** | **Justification** |
| Tree map:  Fig 3.1 – Tree Maps for Attrition by Employee History with Company | -Show the attrition by employee history with company (years at company, years since last promotion and years with current manager)  -Employees with relatively short history with the company (0 – 2 years) are significant contributors, and many of them are from R&D department | Tree map is a great choice when the number of categories of attribute is large, and for displaying large amount of hierarchical data  In Fig 3.1, since there are many values for years, and it is intended to show the proportion of each department associated with them, it is appropriate to use tree map for the visualization. |
| Tree map:  Fig 3.2 – Tree Maps for Attrition by Employee Relationships | -Show the attrition by employee relationships (environment satisfaction, relationship satisfaction and job satisfaction)  -Employees with low environment satisfaction are more likely to quit. However, it is also found that employees may still quit even with high level of relationships and job satisfactions. Again R&D department is the main contributor | Tree map is a great choice when the number of categories of attribute is large, and for displaying large amount of hierarchical data  In Fig 3.2, since the 4 responses (low, medium, high, very high) will be further divided based on department (R&D, HR and Sales) and this introduce some additional hierarchy, hence it is more appropriate to display the visualization using treemap |

Page 4: Attrition by Income, Performance and Salary Hike



|  |  |  |
| --- | --- | --- |
| **Visuals** | **Outcome** | **Justification** |
| Line and stacked column chart:  Fig 4.1 – Attrition and Average of MonthlyIncome by JobRole / JobLevel and Department | -Show the attrition and average of monthly income by job role / job level  -Highest attrition is from job role Laboratory Technician from R&D department (62 no.), with average monthly income of 2.9k, which is among the lowest  -Out of these Laboratory Technician, majority of them are entry level (56 of 62 no.) | For quick comparison of 2 sets of attributes, line and stacked column chart is a solid choice since the 2 visuals (line and bars) are combined together in the same chart  In Fig 4.1, by using line and stacked column chart, we can quickly visualize the attrition not only by job role / job level, but also by additional attribute average monthly income of each job role/ job level, all in the same chart |
| Stacked column chart:  Fig 4.2 – Attrition by JobRole and Performance Rating | -Show the attrition by job role and performance rating  -No clear indication that attrition is related to performance since all staffs received either Excellent or Outstanding for their performance ratings | Stacked column chart work well when the focus of the chart is to compare the totals and one part of the totals  In Fig 4.2, it is used for visualising the attrition by performance ratings across all job roles |
| Line and stacked column chart:  Fig 4.3 – Attrition and Average of PercentSalaryHike by JobRole / JobLevel | -Show the attrition and average of percent salary hike by job role / job level  -Highest attrition job role Laboratory Technician has an average salary hike of 15.4 percent (among the highest)  -Those higher-level job roles (Manager, Director) with relatively lower salary hike are of low attrition | For quick comparison of 2 sets of attributes, line and stacked column chart is a solid choice since the 2 visuals (line and bars) are combined together in the same chart  In Fig 4.3, by using line and stacked column chart, we can quickly visualize the attrition not only by job role / job level, but also by additional attribute percent salary hike of each job role/ job level, all in the same chart |

Page 5: Attrition by Distance from Home and Business Travel



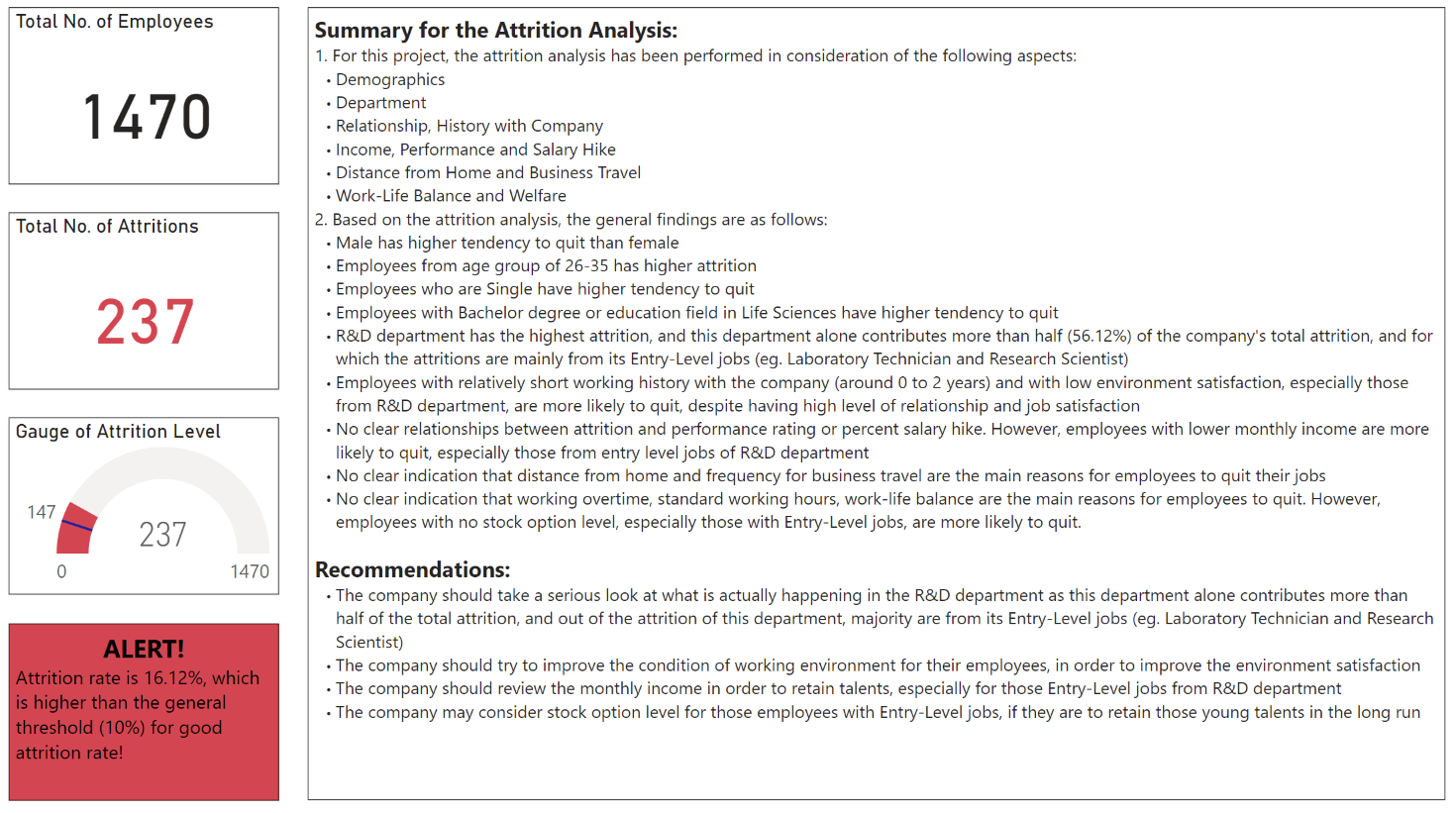
|  |  |  |
| --- | --- | --- |
| **Visuals** | **Outcome** | **Justification** |
| Tree map:  Fig 5.1 – Attrition by DistanceFromHome and JobRole | -Show the attrition by distance from home and job role  -No indication that distance from home is a main factor for employees to quit their jobs | Tree map is a great choice when the number of categories of attribute is large, and for displaying large amount of hierarchical data  In Fig 5.1, since there are many values for distance from home, and it is intended to show the proportion of each job role associated with them, it is appropriate to use tree map for the visualization |
| Tree map:  Fig 5.2 – Attrition by BusinessTravel and JobRole | -Show the attrition by business travel and job role  -More than half of the attrition is contributed by those who rarely travel. Hence, frequency for business travel is not the main reason for employees to quit their jobs | Tree map is a great choice when the number of categories of attribute is large, and for displaying large amount of hierarchical data  In Fig 5.2, since the 3 business travel categories (travel rarely, travel frequently and non-travel) will be further divided based on job roles, and this introduces some additional hierarchy, hence it is more appropriate to display the visualization using treemap |

Page 6: Attrition by Work-Life Balance and Welfare



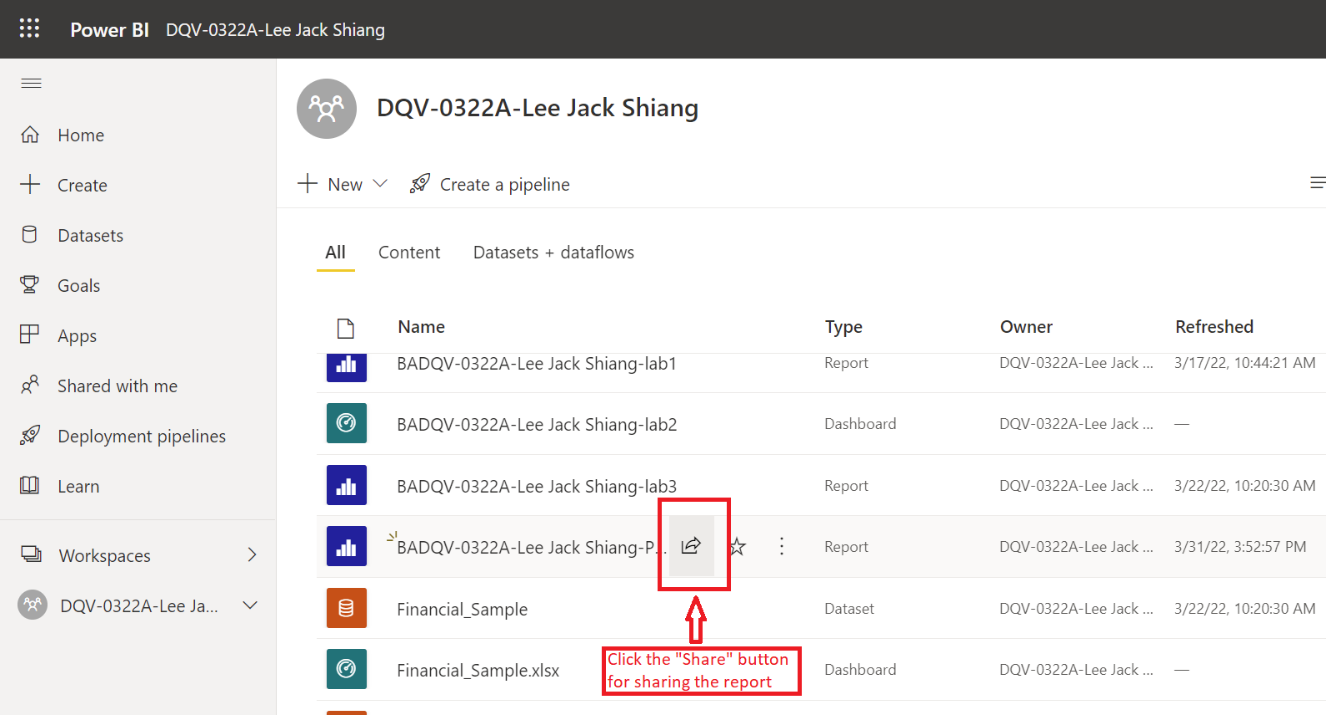
|  |  |  |
| --- | --- | --- |
| **Visuals** | **Outcome** | **Justification** |
| Line and stacked column chart:  Fig 6.1 – Attrition and Average of StandardHours by JobRole / JobLevel and OverTime | -Show the attrition and average of standard hours by job role / job level and overtime  -No clear relationship between attrition and overtime for all job roles since the attrition for each job role is almost evenly distributed (overtime or non-overtime), except for Research Scientist and Manager  -Proportion of working overtime for entry level jobs is slightly higher  -No relationship between attrition and standard hours | For quick comparison of 2 sets of attributes, line and stacked column chart is a solid choice since the 2 visuals (line and bars) are combined together in the same chart  In Fig 6.1, by using line and stacked column chart, we can quickly visualize how the attrition is distributed by overtime for each job role / job level, as well as visualise the average standard hours for all job roles / job levels |
| 100% Stacked column chart:  Fig 6.2 – Attrition by JobRole and WorkLifeBalance | -Show the attrition by job role and work life balance  -No indication that work-life balance is a main factor for attrition since the responses are generally favourable (Good, Better and Best made up majority of the responses) | It is a great choice for showing the relative percentage of multiple categories in stacked columns where the total of stacked columns always equals 100%  In Fig 6.2, since the attrition by work-life balance for each job role is presented in 100% stacked column chart, we can quickly get a sense of how the employees from each job role feel about their work-life balance |
| Clustered column chart:  Fig 6.3 – Attrition by StockOptionLevel and JobLevel | -Show the attrition by stock option level and job level  -Employees with no stock option level are more likely to quit, especially those with entry level jobs | Clustered column chart gives direct comparison of multiple subcategories in a given category and it is a good option for visualising the proportion of each subcategory compared to others  In Fig 6.3, attritions by subcategories (job level) are shown for each category (stock option level). By using clustered column chart, it gives us a quick sense of which stock option level and job level has the highest attrition |

Page 7: Summary and Recommendations



Steps for sharing the report:

1. Click the “Share” button next to the report file in the workspace:



1. In the pop-up window after clicking the “Share” button, the report can be shared via email, Teams, link etc. with related parties:

Graphical user interface, application, Teams

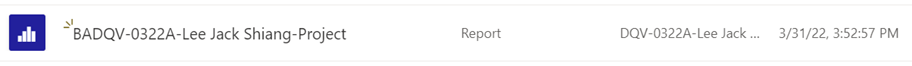
Description automatically generated

1. Annexure (Excel Workbook)

Links for project report, dataset workbook and data definition file for this project:

1. Project report:

<https://app.powerbi.com/groups/6958bd14-c9d9-4bdf-956d-d5942792c45e/reports/878bd80f-dc0b-422f-b31e-3b71bbcc0746/ReportSection>



1. Dataset workbook:

<https://app.powerbi.com/groups/6958bd14-c9d9-4bdf-956d-d5942792c45e/workbooks/1_38585_837227/new>



1. Data definition file:

<https://app.powerbi.com/groups/6958bd14-c9d9-4bdf-956d-d5942792c45e/workbooks/1_38586_837231/new>

